Manuscript: NHESS-2013-202

Memo: response to the reviewers' comments concerning "Predicting the hurricane damage ratio of commercial buildings by claim payout from Hurricane Ike"

Issue	Response/Correction
Reviewer #1	•
The manuscript by Kim et al. describes a statistical analysis to uncover which indicators can be used to predict hurricane damage. The manuscript adds to the current stock of literature on the subject in that it combines various indicators found by other studies to be significantly related to hurricane damage into a single regression analysis (though this could be stressed better in the manuscript). The methodology used in the paper is quite straightforward. Overall, the manuscript has some potential, but definitely needs a bit more body before it can be considered publishing in NHESS.	The authors would like to first thank the editor who allowed us opportunities to revise and resubmit the paper. We also sincerely appreciate the reviewer who provided thorough reviews and valuable comments to help us improve the manuscript. We strongly believe that in the revision we have fully addressed all reviewer's comments and concerns and carefully revised the manuscript based on the feedback we have received. Please see the following sections below responding to the reviewer's comments. Thank you.
p. 3450, l. 22: statement about rising of occurrence of natural disasters due to climate change and abnormal weather needs a reference (especially as it's quite a strong statement, there is a lot of discussion on this subject)	As the reviewer points out, the sentence was revised because there is a lot of discussion on this subject. "The occurrence of natural disasters has been rising exponentially in the United States (Cutter and Emrich, 2005)" Reference Cutter, S. L., and Emrich, C. T. (2005). Are natural hazards and disaster losses in the US increasing?, National Emergency Training Center.
p. 3450, l. 24: there are for sure more recent refs than the 1998 ref stated here. Take a look at the work of L.M. Bouwer.	As the reviewer points out, one reference was added on the sentence. "population explosions in seaside provinces and the sudden expansion of citieshas magnified the risk in those areas (Pielke and Landsea, 1998; Koks et al., 2012)" Reference Koks, E. E., Moel, H., and Bouwer, L. M. (2012). Effect of spatial adaptation measures on flood risk in the coastal area of Flanders, IVM Institute for Environmental Studies.
p. 3450, l. 25: tsunami's are not meteorological disasters as stated here.	As the reviewer points out, tsunami's are not meteorological disasters. The sentence was revised. "meteorological disasters, such as tsunamis, cyclones, deluges, and hurricanes, impact our communities more frequently and critically than any other kind of natural disaster (Cutter and Emrich, 2005)"
p. 3451, l. 1-3: again, this statement needs a reference	As the reviewer points out, one reference was added on the sentence. "Moreover, among the meteorological disasters,

p. 3451, l. 7-8: establishing a model doesn't reduce financial loss at all, it only reveals the potential loss. Actual measures are necessary to reduce the real loss	hurricanes are the most critical and cause the most losses to humankind; therefore, studying hurricanes is crucial in predicting natural disaster damage (Cutter and Emrich, 2005)" Reference Cutter, S. L., and Emrich, C. T. (2005). Are natural hazards and disaster losses in the US increasing?, National Emergency Training Center. Although some damage is unavoidable, establishing a hurricane damage prediction model provides a way to reduce some of the financial loss. Therefore, the establishment of a hurricane damage prediction model provides a way to reveals the potential financial loss.
p. 3451, l. 12: more recent reference necessary as a lot has happened in 18 years (there are various recent studies to hurricanes and damage in NYC for instance)	As the reviewer points out, two references were added on the sentence. "Predicting hurricane damage is a complicated issue, because there is a lack of dependable data and appropriate analyzing methods (Boissonnade and Ulrich, 1995; Colle et al., 2008; Lin et al., 2010)" Reference Colle, B. A., Buonaiuto, F., Bowman, M. J., Wilson, R. E., Flood, R., Hunter, R., Mintz, A., and Hill, E. (2008), New York City's vulnerability to coastal flooding, Bulletin of the American Meteorological Society, 89, 829-841. Lin, N., Emanuel, K. A., Smith, J. A., Vanmarcke, E. (2010), Risk assessment of hurricane storm surge for New York City, 115(18). 112-123.
p. 3452, l. 2: change 'Initially' to 'First'	Done.
p. 3452, l. 8-11: basically says the same as the last sentence before.	As the reviewer points out, the sentence was removed. "To analyze the data, we used a multiple linear regression method to make a global equation, which helps to identify the relationship between the dependent variable and independent variables. Utilizing the statistical method, this research then identified the relationship between TWIA claim payouts and the vulnerability indicators."
p. 3452. l. 17: add 'At the time' before 'Hurricane Ike' as with Sandy in 2012 I think Ike moved down to fourth place now.	Refer to the National Hurricane Center, Hurricane Ike was the third most costly hurricane to hit the United States after hurricanes Katrina and Sandy. The rank is (cost refers to total estimated property damage); 1. Hurricane Katrina 2. Hurricane Sandy

	3. Hurricane Ike
	4. Hurricane Andrew
	Therefore, the sentence was revised as below: "Hurricane Ike was the third most costly hurricane to hit the United States after hurricanes Katrina and Sandy"
p. 3452, general: some words are necessary on how Ike caused damage. This can differ substantially between hurricane. Take for instance Andrew, which caused mainly storm damage, and Katrina, which mainly caused damage by flooding. Also the type of damage covered by TWA and how this relates to the damaging effect of Ike is crucial.	The damage records from TWIA were mixed the storm and flood damages. That is the reason we counted the damages as one. Hence, we do not know about the details of the damages.
Does the dataset used in this study only relate to storm losses? (as I would infer from line 25)	
p. 3453, l. 5: add 'spatial' before 'distribution'	Done
p. 3453, l. 7-8: why do you use claims from a duration of 3.5 years? And more strikingly: you use claims dating from a month before the actual storm hit?	We also wonder the points and asked the manager and staffs of TWIA. The answer was that the claims are usually take a long to get it. Because they have to exam the building and some claims are related legal issues. Hence, some of the claims took several years. The claims dating from a month before the actual storm hit would be recorded by mistakes. But we sure about that the
	records are all about Hurricane Ike.
p. 3453, l. 9-18: this whole paragraph just sums up the numbers found in the table, so adds nothing. Highlighting the important message from the table is sufficient. Also:	As the reviewer points out, the sentence was removed and Table 1 was revised.
too many significant numbers are used, i.e. 'overall claim payout of 450 M\$'	"As shown in Table 1, the damages were happened through Texas coastal counties. Galveston, Jefferson, Brazoria, and Chamber had most damage from Hurricane Ike. Especially, Galveston was most damaged in terms of the number of claims and the dollar amount of damage"
	There were a total of 4150 claims, with an overall claim payout of \$450 518 330. The most damaged county was Galveston, both in terms of the number of claims (1807, 43.54% of the total number of claims) and the dollar amount of damage (\$255 333 818, 56.68% of the total dollar amount). The other damaged counties in Texas were: Jefferson County with \$104 249 917 in total losses and 1218 claims, Brazoria County with \$46 922 396 in total losses and 597 claims, Chambers County with \$39 755 609 in total losses and 470 claims, Harris County with \$4 126

	821 in total losse \$36 981 in total l with \$67 501 in t County with \$52 Table 1. TWIA cl County Galveston Jefferson Brazoria Chambers Harris Matagorda Liberty Nueces Total	osses and nin otal losses an 87 in total los	e claims ed two classes and the county	, Liberty C aims, and N	ounty Nueces
p. 3453, l.19-22: why go through the trouble of getting a random sample of 500 instead of using all your data? 4150 data points should be no problem for a linear regression. Also an example is given for a sample population of 5000, whilst the real population is 4150. p. 3454, l. 14: what do you mean with 'wind analysis'? p. 3454, l. 15-16: you're describing data, not an analysis.	We assumed that think the number assumption of state. The reason is that enough sample size the enough samples (500). We as reviewer point " This analysis"	of sample are tistics. when the sample is 370. Hence size is smalled believe that the sample believe that the sample out, the seron data consists	nple popu ce, the re er than 3 we had e e gauged intence w	lation is 50 al population of the nough sample data'. as revised.	o0, the on is 4150, had more ple size.
p. 3454, l. 17: I fail to see how the swath map gives any indication of damage, extra data (on assets) and a model is necessary for that. Same goes for line 23	As follow the con Line 17. Using the determine the win hurricane damage Line 23. With the their desired area wind and hurricane	mment, ne swath map, nd parameters e with extra d ese data, resea , time, and hu	, investig s but are amage d archers c urricane,	gators can r also able to ata. an create n and can ex	naps for
p. 3453, 1. 26-27: on the left and right side of the hurricane, doesn't this also relate on the direction the system moves? I can understand this distinction in a system that moves from east to west, but what if a systems moves from south to north?	The means is that hurricane moves for are located on the have less damage a hurricane path.	Forward from see west side of	south to r	orth. Prope ane path ty	erties that pically

p. 3455, l.4: robust and variable winds? Those characteristics sound quite conflicting to me.	Which means that "strong"
p. 3455, l. 4-7: it would be nicer to let the analysis show if this is indeed the case! That's exactly what you can use correlation and regression analysis for.	As reviewer points out, the sentence was revised. In the Northern Hemisphere, the left side of a hurricane path generally creates less strong winds, which makes properties on this side less vulnerable to hurricane damage. "The two different actions of hurricanes, counterclockwise rotation and forward movement, are combined in the right side of hurricanes and then the right side has broader and stronger winds. Conversely, properties on the left side of a hurricane path are less prone to losses"
p. 3455, l. 11: to assess the probability of loss the insurer would also have to know about the frequency of events, when evaluating the insured built environment one can only say something about the a 'probable' or 'possible' loss (not the probability) p. 3455, l. 12-13: the vulnerability is not influenced by the magnitude of the loss, but the magnitude of the loss is determined by	As reviewer points out, the sentence was revised. "The insurer should evaluate the insured built environment to measure the vulnerability in order to assess the probability of possible loss" The intensity of exposure to natural disasters and the magnitude of loss is determined by the vulnerability of a built environment (Khanduri and Morrow, 2003).
p. 3455, end: what data is used for the area and age of buildings? Is this the year of building, or the year of last substantial renovation? Also: why not try more indicators and see if they maybe affect hurricane damage as well? (like amount of floors, foundation type, building material, etc.)	We collected the data from (area and age of buildings) each website of the appraisal district for each Texas coastal county, based on a 2008 roll. In addition, we used actual year built for counting building age. We also try to collect more indicator you mentioned, However, unfortunately, Galveston appraisal district only have the detailed building information. (like amount of floors, foundation type, building material, etc.). Other counties have just basic information (Age, area, etc.) We will study hurricane damage in Galveston with the detailed building information in the following paper.
p. 3456, l. 27: I'd consider rephrasing those final sentences saying ' should be considered when prediction hurricane damage' into 'In this study we thus consider for predicting hurricane damage'. (also goes for earlier instances)	As reviewer points out, the sentence was revised. As geographical vulnerability indicators, FEMA Flood Zones, Hurricane Surge Zones, and distance from water should be considered when predicting hurricane damage. "In this study we thus consider, FEMA Flood Zones, Hurricane Surge Zones, and distance from water for

	predicting hurricane damage"
p. 3457, l. 6: unconditional financial damage?	changed "The purpose of this model is to predict the percentage of damages in the building properties"
p. 3457, 6-9: explain why this ratio is chosen as indicator, and what it means (the percentage of property damaged?)	We believe that it should be worth it to measure the percentage of damages in the building properties. Due to the reason, we used the ratio as an indicator. We will also study only TWIA claim payout as an indicator to measure the financial loss in building property in following paper.
p. 3457, l.10: I don't see components 5, 6 and 7 in the equation, though they are explained	$\begin{aligned} &Ratio = \beta_0 + \ \beta_1 \cdot Wind_Speed + \ \beta_2 \cdot Side_Right + \ \beta_3 \cdot Age + \ \beta_4 \\ &\cdot Area + \ \beta_5 \cdot FEMA_Zones + \ \beta_6 \cdot Surge_Zones + \ \beta_7 \\ &\cdot Dist_Shore \end{aligned}$
p. 3457, l. 19: what is the difference between X-zone and unregistered?	FEMA flood zone X have a 0.2%, or smaller, chance to flood on any given year. Unregistered zone is the out of the FEMA flood zones.
p. 3457, l. 22: I woud call this section differently: Results.	As reviewer points out, the section title was changed to "Results".
p. 3458, l. 19-20: this sentence adds nothing. Discuss your results here: 'all coefficients have positive values, indicating ' (same	p. 3458 line19-20
goes for final sentence of 4.2)	The coefficients imply the linear relationship within a scale of -1 to +1, and the sign of the coefficients define whether the correlation is negative or positive.
	of -1 to +1, and the sign of the coefficients define whether
	of—1 to +1, and the sign of the coefficients define whether the correlation is negative or positive. "The max. wind speed and building age have positive sign of the coefficients. It defines the indicators have positive correlation with ratio. On the other hand, the building area and distance from the property centroid to shoreline have negative sign of the coefficients. It indicates the

	Also, Table 5 caption was revised:
	Table 5. Results of Spearman Correlation Analysis for Ordinal Variables used in Regression
p. 3459, l. 23-26: this part looks straight from a textbook and says nothing about your results. Instead of this part discuss what your significant relationship actually means.	As reviewer points out, the sentence was removed and the results were added. The model is statistically significant because the calculated
	P value of 0.000 is less than 0.05. This means that there is a significant linear relationship between the dependent variable and the independent variables. The null hypothesis, which states that there is no linear relationship between the dependent variable and the independent variables, should be rejected.
	"The model is statistically significant, which means there is a linear relationship between the dependent variable and the independent variables"
p. 3460, l. 15-16: why 'scaled from 0 to 1', standardization already allows to compare variables with different units.	The meaning of the sentence is that the variance of the standardization is 1.
	"The standardized coefficients, also called beta coefficients, were scaled from 0 to 1 and then employed to reveal which independent variables had more effect on the ratio when the variables are various units"
p. 3460, l. 24: \$=\$ should be \$/\$? (or just	As reviewer points out, We removed it.
remove it)	Log(Predicted Ratio (\$/\$))
p. 3460, 1.2 vs. 1. 22: there seems to be an inconsistency in the text. First 33.7% of the variability is explained and later 34.3%. In any case, I would not use decimals, two significant numbers is enough.	These were revised to 34%.
p. 3460, l. 6-9: other factors were found significant as well (see Table 4 and 5), like wind speed. But because of strong autocorrelation between variables (wind speed and distance to shoreline in this case), one of the two is enough for the model.	The ratio and wind speed have a positive correlation in Table 4. However, in the statistically model, the wind speed is not significant. The goals of study is to make a best damage function. Hence, we rejected the wind speed in the model.
Please explain such details in your results.	We conducted the Durbin-Waton test for checking the autocorrelation of the model. The value was 0.033. Therefore, we would believe that there is no autocorrelation in the model.
p. 3461, section 4.4: can be removed (is same as 4.3)	As reviewer points out, the section was removed with Fig.9.
	4.4 Statistic model and validity In the ratio regression, four indicators were proven to be

	significant predicators for the transformed ratio. The scale of the Variance Inflation Factor (VIF), from 1.022 to 2.180, verified that there is no multicollinearity among the independent variables, which proved that the independent variables are not correlated. The adjusted R²-value of the model is 0.337; therefore, the transformed ratio is able to describe with 33.7% of variability in the data with the four significant predictors. The scatter plot of the actual log transformed ratio versus the predicted log ratio is depicted in Fig. 9.
p. 3461, l. 17: I see no figure of a map with the spatial distribution of losses (which could easily be done with the formula that is calculated), only of the payouts (which is input for the study, not so much a result).	This study identified the vulnerability predictors for hurricanes, establishing a metric to predict the financial losses from hurricanes, and created a map showing the spatial distribution of the loss and vulnerabilities to identify hurricane prone areas.
	This study identified the vulnerability predictors for hurricanes, establishing a metric to predict the financial losses from hurricanes.
p. 3461, l. 20-21: the ratio (\$/\$) should be unit-less I'd say, so the regression does not say anything about real pecuniary losses, but about the share of total possible loss.	As the dependent variable, we used the ratio of the value of the Texas Windstorm Insurance Association's (TWIA) claim payout divided by the appraised values of the buildings to predict the share of total possible losses, and to find significant predictors.
p. 3462, l.10-14: remove this part, focus on the findings of your analysis. How could the TWA use your results? I.e. what use is (spatial explicit) hurricane prediction for the TWA?	As reviewer points out, the sentence was removed. The overall number of claims was 4150, and the overall claim payout amount was \$450 518 330. The county that suffered the most damage was Galveston, both in terms of the number of claims, (1807, 43.54 %) and the dollar amount of damage (\$255 333 818, 56.68 %). Thus, the damage distributions verify that Galveston county is the most vulnerable to hurricanes in Texas.
p. 3462, l. 15-16: avoid this kind of textbook sentences	As reviewer points out, the sentence was revised. The ratio statistic model is significant because the calculated P value of 0.000 was less than 0.05. This proves that the independent variables are able to predict the ratio. "The ratio model is statistically significant. This proves that the independent variables are able to predict the ratio"
Tables and Figures:	

A table with the independent variables used in the study, their sources (of data), unit, and studies which have studied them before would give a good overview for the reader.	Done.
Table 3: the median and 50 percentile are given, which are exactly the same, so one can be removed from the table.	Done.
Table 4, caption: add: ': for continuous variables used in regression'. Add mark for significance (possibly even for different levels).	As reviewer points out, the caption was revised to "Results of Pearson correlation analysis for continuous variable used in regression"
Table 5, caption: should be Results of Spearman correlation analysis. And I would add ' for ordinal variables used in regression'.	As reviewer points out, the caption was revised to "Results of Spearman correlation analysis for ordinal variables used in regression"
I would merge Table 6 and 7 to save space.	Done.
Table 9: show would be nicer to show all coefficients of the regression model, so the reader can see for him/herself which ones are significant.	Done.
Figure 1: shows not only data collection, but also the analysis (i.e. regression model is not data collection)	As reviewer points out, the caption was revised to "Research methodology"
Figure 3: the legend seems off as it shows standard deviation. Should be something related to wind speed I presume	Done.
Figure 4: can be removed, also doesn't match the caption at all	As reviewer points out, this figure was removed.

Issue	Response/Correction
Reviewer #2	
Referee #1 has already developed extensive general and detailed comments, to which I am totally supportive (general and detailed comments). In order provide hopefully some additional help for the authors to revise appropriately their paper, which has indeed a potential for publication after revision, I shall just summarize my general impression on the version submitted, in the following 3 main remarks:	The authors would like to first thank the editor who allowed us opportunities to revise and resubmit the paper. We also sincerely appreciate the reviewer who provided thorough reviews and valuable comments to help us improve the manuscript. We strongly believe that in the revision we have fully addressed all reviewer's comments and concerns and carefully revised the manuscript based on the feedback we have received. Please see the following sections below responding to the reviewer's comments. Thank you.
1. As insurance claims payouts data have been mobilized, this interesting raw material to work on should be more thoroughly described/discussed in its contents (how are they defined? for which assets / insurance coverage?), limits of professional use, for reader to get a clearer picture of the context and the potential of the exercise. The uncertainties about the geocoding / GIS treatment of the claims data should be more addressed/discussed: which categories of coordinates? which grid?	It was our hope that investigates the damages on the all type of buildings. However, we are only allowed to get the limited payout data (commercial buildings) due to the policy of the TWIA. Therefore, private properties were not addressed in this study. That is the one of the reasons that investigates the property of commercial buildings. The raw data is included; street address (number, street, city, zip code), commercial property damage loss(\$) (the TWIA payout associated with hurricane Ike), TWIA payout date (the date TWIA paid for the property damage loss). We did not know about the insurance coverage. We used ArcGIS address locator for geocoding. The claim payouts were plotted on each county parcel of the shape files by using the ArcGIS address locator. The Geographic Coordinate System was GCS_North_American_1983 and Datum was D_North_American_1983. We added some sentence on the paper. Page 3452. Line 1. " This research was conducted as described in the following process (Figure 1). First, we used the ArcGIS address locator to overlap the TWIA claim payout properties onto the study areas. The Geographic Coordinate System was GCS_North_American_1983 and the Datum was D_North_American_1983" Page 3453. Line 3. " The observational units in this research are the insured claim payouts from TWIA, of the appraised commercial buildings hit by Hurricane Ike. The raw data was included;

street address (number, street, city, zip code), commercial property damage loss(\$) (the TWIA payout associated with hurricane Ike), TWIA payout date (the date TWIA paid for the property damage loss). Private properties were not included due to the policy of the TWIA..."

2. Appropriate developments are missing on the general intention and practical potential use of the conceptual model introduced by the paper for respectively the (re)insurance industry and public authorities, mainly: common goals and respective specific issues for downstream processing?

It is not clear if this study ends up "to establish a metric to predict the financial losses of huricanes". A parallel might be drawn between this correlation study on relevant risk factors and the design research required for the different modules which build a loss modeling tool, in particular the damage functions part of it: as the set of data seem to provide satisfactory statistic correlations between asset damage ratios, with vulnerability factors such as building age and hazard indicators such as the wind speed (flood height?) at (or near) the location of the asset (or on a surface basis?). comments/conclusions would be welcome on the damage functions to be retained and their limits of confidence?

on the damage functions to be retained and their limits of confidence?

Does this study teach us something to reduce uncertainties of the modeling tools, depending to their context of use to predict the financial losses of hurricanes (commercial or research)?

As reviewer points out, some parts were revised.

Page 3453. Line 24.

"... This research used the Texas Windstorm Insurance Association (TWIA) claim payout records of commercial buildings from Hurricane Ike to identify hurricane and vulnerability predictors, establish a metric to predict the financial losses of hurricanes, and image the spatial distribution of the loss and vulnerabilities to identify hurricane-prone areas. This damage function will determine if the developed models are verifiable; additionally, this function will calculate the significant relationships among economic losses (i.e., insured loss payments), vulnerability indicators, and hurricane indicators. This model and findings may together become one of the most useful and vital references for hurricane damage prediction for public works, as well as other entities such as government agencies, emergency planners, and insurance companies. For instance, insurance companies may be able to adjust their policies to follow the indicators, and therefore enjoy more profit. This model should become an important guideline to be used by government agencies and local emergency planners who need to identify the exact relationship between hurricanes and vulnerability indicators...."

Page 3463. Line 24.

"...Moreover, damage function might have reduce uncertainties of the modeling tools, since we statistically investigated real damage records. However, the damage function would be limited in mega hurricane like Hurricane Ike. The reason is that we investigated only a mega hurricane..."

3. As already stated by Referee #1 and following previous remarks, the summary and conclusions should be implemented/reformulated, for this paper to emphasize on the innovative added values of the work carried out as an applied research on risk evaluation/prediction, instead of giving the disappointing impression of a statistical study report (with still weaknesses in the way some intermediary figures and results are displayed).

As reviewer points out, conclusion section was revised. In addition, discussion section was added.

5. Discussion

"...This research used the appraised commercial building's claim payouts from the Texas Windstorm Insurance Association (TWIA) for damages caused by Hurricane Ike in Texas. The range of the observational unit was from 17 August 2008 to 22 February 2012. The ratio model is statistically significant. This proves that the independent variables are able to predict the ratio. The adjusted R² value of 0.337 indicates that 33.7% of the variability in the transformed ratio can be described by the significant predictors...."

6. Conclusions

"...The developed statistical model and results form an important guideline for insurance companies and emergency planners when predicting hurricane damage. For instance, following our indicators, insurance companies can adjust and reconsider their policies for increased profits. Using our model, government agencies and emergency planners can identify hurricanes and the built environment and geographic vulnerability indicators, and then evaluate the effects of each factor with respect to hurricane risk for improved hurricane damage predictions. Through developed statistical models, it is possible that other states may at some point be able to identify the significant relationships among the indicators in order to assess their own possible hurricane losses. The vulnerability indicators included in this study will help to identify building environment and geographic vulnerabilities, as well as evaluate the effect of each factor with respect to damage from hurricanes in order to mitigate perceived danger. Additionally, the significant hurricane indicators will help to improve hurricane damage prediction and also would help to build other damage functions by the indicators. Moreover, the damage function might have reduce uncertainties of the modeling tools, since we statistically investigated real damage records. However, the damage function would be limited in mega hurricane like Hurricane Ike. The reason is that we investigated only a mega hurricane..."